

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:31 ; Search time 91.75 seconds  
(without alignments)  
22.175 Million cell updates/sec

Title: US-09-331-631a-1\_COPY\_186\_248

Perfect score: 335  
Sequence: 1 KRDPQREYEDCDRCRRCEQOE.....MMNPQGGSGRYEGEEGEOQS 63

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 segs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114	32.1	588	1	VCIA_GOSHI
2	101	28.5	605	1	VCIA_GOSHI
3	84	23.7	966	1	SSN6_YEAST
4	79	22.3	919	1	ANDR_HUMAN
5	75	21.1	1090	1	NIT4_NEUCR
6	74	20.8	1898	1	TRHY_HUMAN
7	74	20.8	2442	1	CBP_HUMAN
8	73.5	20.7	185	1	T2_MOUSE
9	73	20.6	573	1	GLB1_MAIZE
10	73	20.6	911	1	ANDR_PANTR
11	70.5	19.9	1344	1	IF3A_MOUSE
12	69	19.4	239	1	WT1_SMIMA
13	69	19.4	288	1	WT1_ALLMT
14	69	19.4	448	1	WT1_RAT
15	69	19.4	449	1	WT1_HUMAN
16	69	19.4	449	1	WT1_MOUSE
17	69	19.4	449	1	WT1_PIG
18	69	19.4	895	1	ANDR_PAPHA
19	69	19.4	1023	1	GLT_DROME
20	68.5	19.3	758	1	YH38_YEAST
21	68.5	19.3	1382	1	IF3A_HUMAN
22	68	19.2	1154	1	WC1_NEUCR
23	68	19.2	1265	1	CYAS_CANFA
24	67.5	19.0	164	1	2SS2_ARATH
25	66.5	18.7	170	1	2SS2_ARATH
26	66.5	18.7	372	1	HXA2_MOUSE
27	66.5	18.7	372	1	HXA2_RAT
28	66	18.6	423	1	TF3A_HUMAN
29	66	18.6	529	1	GAG_MLVDU
30	66	18.6	536	1	GAG_MLVDE
31	66	18.6	538	1	GAG_MLVFP
32	66	18.6	539	1	GAG_MLVFS
33	66	18.6	540	1	GAG_MLVVO

34	66	18.6	614	1	RU17_HUMAN
35	66	18.6	1407	1	TRHY_RABIT
36	66	18.6	1596	1	MAM_DROME
37	65.5	18.5	47	1	AGRP_LUECY
38	65.5	18.5	255	1	LP61_EIMTE
39	65.5	18.5	450	1	SRF_DROME
40	65.5	18.5	568	1	INVO_RAT
41	65.5	18.5	2124	1	Y192_HUMAN
42	65	18.3	280	1	YT24_AGRVI
43	65	18.3	339	1	TF2D_HUMAN
44	65	18.3	384	1	INVO_GALCR
45	65	18.3	905	1	SNFS_YEAST

## ALIGNMENTS

RESULT	ID	VCIA_GOSHI	STANDARD	PRT	588 AA
AC	P09801				
DT	01-MAR-1989	(Rel. 10, last sequence update)			
DT	01-MAR-1989	(Rel. 10, last sequence update)			
DT	15-JUL-1999	(Rel. 38, last annotation update)			
DE	VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B)				
OS	Gossypium hirsutum (Upland cotton)				
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RA	Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;				
RT	"Developmental biochemistry of cottonseed embryogenesis and germination. XVII. cDNA and amino acid sequences of the members of the storage protein families."				
RL	Plant Mol. Biol. 7:475-489(1986).				
CC	-1- FUNCTION: SEED STORAGE PROTEIN.				
CC	-1- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN BODIES.				
CC	-1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN, CONVICILIN, CONGLICININ, ETC.).				
CC					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	EMBL: M16891; AAA33071.1; -				
DR	PIR: A30838; FWCNAB.				
DR	HSSP: P50477; ICAX.				
DR	INTERPRO: IPR001113; -				
DR	PFAM: PF00546; Seedstore-7s; 1.				
FT	Seed storage protein; Signal.				
FT	SIGNAL 1 25				
FT	CHAIN 26 588				
SQ	SEQUENCE 588 AA; 69729 MW; 636699B29AB8ADEB CRC64;				

Query Match 32.1%; Score 114; DB 1; Length 588;

Best local similarity 33.7%; Pred. No. 9.6e-05; Matches 33; Conservative 9; Mismatches 18; Indels 38; Gaps 5;

QY 3 DPQREYEDCDRCRRCEQOEPRQHQQLRC-----REQ-----ROHGRGDMM---45.

DB 82 DPQR- YEECDQECRQOERORRPOCOQRLKRPEDQOQSOFQRCQGHCHQOQERPR 140

QY 46 -----NPR-----GSGRYEGEEGEOQS 63

DB 141 KOQVRECRERYQENPWRREBEAEAEETEGEEOQS 178

```

RESULT      2
VCIA_GOSHI ID VCIA_GOSHI STANDARD: PRT: 605 AA.
AC P09799;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-JUL-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
DS Gossypium hirsutum (Upland cotton).
OC Eukaryota, Viridiplantae, Embryophyta, Tracheophyta, Spermatophyta;
OC Magnoliophyta, eudicotyledons, core eudicots, Rosidae, eurosids II;
OC Malvales, Malvaceae, Gossypium.
RN [1]
RE SEQUENCE FROM N.A.
RA Chlan C.A., Borroto K., Kamalay J.A., Dure L. III:
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XIX. Sequences and genomic organization of the alpha
RT globulin (vicillin) genes of cottonseed.";
RL Plant Mol. Biol. 9:533-546(1987).
CC -I FUNCTION: SEED STORAGE PROTEIN.
CC -I SUBCELLULAR LOCATION: CYTOSOL;MEMBRANE BOUND VACUOLAR PROTEIN
CC BOBIES.
CC -I SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, COMLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19378; AAA33069.1; -.
DR PIR: S06398; S06398.
DR HSSP: P50477; ICAX.
DR INTERPRO: IPR001113; -.
DR Pfam: PF00546; Seedstore_7s; 1.
KW Seed storage protein; Signal.
FT SIGNAL 1 23
FT CHAIN 24 605 VICILIN GC72-A.
SQ SEQUENCE 605 AA; 71049 MW; C9DB9J7IC976953B CRC64;

Query Match 28.5%; Score 101; DB 1; Length 605;
Best Local Similarity 27.7%; Pred. No. 0.002;
Matches 26; Conservative 15; Mismatches 19; Indels 34; Gaps 4;

QY 3 DPQQREYEDCRRCRCEQPFRQHQCOLCRREQ-QQHGRGGGDM----- 45
||| | |||| :||: | : | : |||| : : |
Db 79 DPQR-RYDQCRRCHCGEERLRPHCEGCSCREGYEKQQQDDQPKQFKRCEQGQNRQMGEQRPE 137
Yy 46 -----NPDRG-GSGRYEEGEGQS 63
: | : : | ||| :

Db 138 RKQCKVECREQYOEDPWKGRENKMREEEES 171

RESULT      3
SSN6_YEAST ID SSN6_YEAST STANDARD: PRT: 966 AA.
AC P14922;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE GLUCOSE REPRESSION MEDIATOR PROTEIN.
DS SSN6 OR CYC8 OR YBR112C OR YRB0908.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota, Fungi, Ascomycota, Saccharomycetes, Saccharomycetales;
OC Saccharomycetaceae, Saccharomyces.
RN [1]
RE SEQUENCE FROM N.A.
RP MEDLINE=892111964; PubMed=2854095;
XX
```

RA Trumbly R. J. ;  
RT "Cloning and characterization of the CYC8 gene mediating glucose  
RT repression in yeast. ;"  
RL Gene 73:97-111(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88065502; PubMed=3316983;  
RA Schultz J., Carlson M. ;  
RT "Molecular analysis of SSN6, a gene functionally related to the SNF1  
RT protein kinase of *Saccharomyces cerevisiae*. ;"  
RL Mol. Cell. Biol. 7:3637-3645(1987).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SS288C;  
RX MEDLINE=92327848; PubMed=1626431;  
RA Mannhaupt G., Stucke R., Ehnlé S., Vetter I., Feldmann H. ;  
RT "Molecular analysis of yeast chromosome II between CMD1 and LYS2: the  
RT excision repair gene RAD16 located in this region belongs to a novel  
RT group of double-finger proteins. ;"  
RL Yeast 8:397-408(1992).  
RN [4]  
RP TPR REPEATS.  
RX MEDLINE=90124639; PubMed=240612;  
RA Sliozeki R.S., Boguski M.S., Goebel M., Hieter P.A. ;  
RT "A repeating amino acid motif in CDC3 defines a family of proteins  
RT and a new relationship among genes required for mitosis and RNA  
RT synthesis. ;"  
RL Cell 60:307-317(1990).  
CC -I- FUNCTION: IT IS INVOLVED IN REPRESSION BY A1-ALPHA2 AND ALPHA2 AND  
CC IN OTHER SYSTEMS AS A GENERAL REPRESSOR OF TRANSCRIPTION. THIS  
CC PROTEIN HAS NO OBVIOUS DNA-BINDING DOMAINS. IT MIGHT NOT INTERACT  
CC DIRECTLY WITH DNA BUT WITH DNA-BOUND PROTEINS.  
CC -I- SUBCELLULAR LOCATION: NUCLEAR.  
CC -I- SIMILARITY: CONTAINS 10 TPR DOMAINS.  
CC -I- SIMILARITY: TO YEAST GAL1 AND CCK4.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M23440; AAA34545.1; -;  
DR EMBL; M17826; AAA35103.1; -;  
DR EMBL; X66247; CAA46973.1; -;  
DR EMBL; X78993; CAA55615.1; -;  
DR EMBL; Z35981; CAA85069.1; -;  
DR PIR; S25365; S25365.  
DR SGD; S0000316; CYC8.  
DR INTERPRO; IPR001440; -;  
DR PFAM; PF00515; TPR; 7.  
KW Transcription regulation; Repressor; Repeat; TPR domain;  
KW Nuclear protein.  
FT DOMAIN 15 30 POLY-GLN.  
FT REPEAT 46 79 TPR 1.  
FT REPEAT 80 113 TPR 2.  
FT REPEAT 114 147 TPR 3.  
FT REPEAT 150 183 TPR 4.  
FT REPEAT 187 220 TPR 5.  
FT REPEAT 224 257 TPR 6.  
FT REPEAT 258 291 TPR 7.  
FT REPEAT 296 329 TPR 8.  
FT REPEAT 330 363 TPR 9.  
FT REPEAT 364 398 TPR 10.  
FT DOMAIN 493 556 30 X 2 AA TANDEM REPEATS OF Q-A.  
FT POLY-GLN.  
FT CONFLICT 547 587 K -> Q. (IN REF. 3).  
SQ SEQUENCE 966 AA; 107202 MW; 84B509C3208C5C0 CRC64;

Query Match

23.78; Score 84; DB 1; Length 966;

Best Local Similarity 34.08; Pred. No. 0.16;  
Matches 17; Conservative 15; Mismatches 18; Indels 0; Gaps 0;  
1 KRDPQREVEDCRRCRCEOPROOHOCLOPCREOQROHGRGGMNORG 50  
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|  
Db 562 000000000000000000000000LQPLPRQOLQKGVSVOMLNPQG 611

RESULT 4  
ANDR\_HUMAN STANDARD; PRT; 919 AA.  
ID P10275;  
AC 01-MAR-1989 (Rel. 10, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ANDROGEN RECEPTOR (DIHYDROTESTOSTERONE RECEPTOR).  
GN AR OR NR3C4 OR DHTR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89112208; PubMed=3216866;  
RA Lubahn D.B., Joseph D.R., Sar M., Tan J., Higgs H.N., Larson R.E.,  
RA French F.S., Wilson E.M.;  
RT "The human androgen receptor: complementary deoxyribonucleic acid  
RT cloning, sequence analysis and gene expression in prostate.";  
RL Mol. Endocrinol. 2:1265-1275(1988).  
[2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90083302; PubMed=2594783;  
RA Lubahn D.B., Brown T.R., Simental J.A., Higgs H.N., Migeon C.J.,  
RA Wilson E.M., French F.S.;  
RT "Sequence of the intron/exon junctions of the coding region of the  
RT human androgen receptor gene and identification of a point mutation  
RT in a family with complete androgen insensitivity.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:9534-9538(1989).  
[3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90258935; PubMed=2342476;  
RA Govindan M.V.;  
RT "Specific region in hormone binding domain is essential for hormone  
RT binding and trans-activation by human androgen receptor.";  
RL Mol. Endocrinol. 4:417-427(1990).  
[4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89017168; PubMed=3174628;  
RA Chang C., Kokontis J., Liao S.;  
RT "Structural analysis of complementary DNA and amino acid sequences of  
RT human and rat androgen receptors.";  
RL Proc. Natl. Acad. Sci. U.S.A. 85:7211-7215(1988).  
[5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90083302; PubMed=2594783;  
RA Lubahn D.B., Joseph D.R., Sar M., Tan J., Higgs H.N., Larson R.E.,  
RA French F.S., Wilson E.M.;  
RT "The human androgen receptor: complementary deoxyribonucleic acid  
RT cloning, sequence analysis and gene expression in prostate.";  
RL Mol. Endocrinol. 2:1265-1275(1988).  
[6]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91155943; PubMed=2293020;  
RA Marceli M., Tilley W.D., Wilson C.M., Griffin J.E., Wilson J.D.,  
RA McPhaul M.J.;  
RT "Definition of the human androgen receptor gene structure permits the  
RT identification of mutations that cause androgen resistance: premature  
RT termination of the receptor protein at amino acid residue 588 causes  
RT complete androgen resistance.";  
RL Mol. Endocrinol. 4:1105-1116(1990).  
[7]  
RP SEQUENCE OF 189-919 FROM N.A.

RX MEDLINE=8817811; PubMed=3353726;  
RA Chang C., Kokontis J., Liao S.;  
RT "Molecular cloning of human and rat complementary DNA encoding  
RT androgen receptors.";  
RL Science 240:324-326(1988).  
[8]  
RP SEQUENCE OF 468-919 FROM N.A.  
RX MEDLINE=88240407; PubMed=337788;  
RA Trapman J., Klaassen P., Kuiper G.G.J.M., van der Korput J.A.G.M.,  
RA Faber P.W., van Rooij H.C.J., Geurts van Kessel A., Voorhorst M.M.,  
RA Mulder E., Brinkmann A.O.;  
RT "Cloning, structure and expression of a cDNA encoding the human  
RT androgen receptor.";  
RL Biochem. Biophys. Res. Commun. 153:241-248(1988).  
[9]  
RP POLYMORPHISM OF POLY-GLN REGION.  
RX MEDLINE=92220629; PubMed=1561105;  
RA Stedens H.F., Oostera B.A., Brinkmann A.O., Trapman J.;  
RT "Trinucleotide repeat polymorphism in the androgen receptor gene  
RT (AR).";  
RL Nucleic Acids Res. 20:1427-1427(1992).  
[10]  
RP POLYMORPHISM OF POLY-GLY REGION.  
RX MEDLINE=91287825; PubMed=2062380;  
RA Spada A.R., Wilson E.M., Lubahn D.B., Harding A.E., Fischback K.H.;  
RT "Androgen receptor gene mutations in X-linked spinal and bulbar  
RT muscular atrophy.";  
RL Nature 352:77-79(1991).  
[11]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=95023089; PubMed=7937057;  
RA Patterson M.N., Hughes L.A., Gottlieb B., Pinsky L.;  
RT "The androgen receptor gene mutations database.";  
RL Nucleic Acids Res. 22:3560-3562(1994).  
[12]  
RP REVIEW ON VARIANTS.  
RX MEDLINE=97169385; PubMed=9016528;  
RA Gottlieb B., Trifiro M., Lombroso R., Vasilion D.M., Pinsky L.;  
RT "The androgen receptor gene mutations database.";  
RL Nucleic Acids Res. 25:158-162(1997).  
[13]  
RP VARIANT LNCAP ALA-877.  
RX MEDLINE=91083633; PubMed=2260966;  
RA Veldscholte J., Ris-Stalpers C., Kuiper G.G., Jenster G.,  
RA Berrevoets C., Klaassen E., van Rooij H.C.J., Trapman J.,  
RA Brinkmann A.O., Mulder E.;  
RT "A mutation in the ligand binding domain of the androgen receptor of  
RT human LNCap cells affects steroid binding characteristics and  
RT response to anti-androgens.";  
RL Biochem. Biophys. Res. Commun. 173:534-540(1990).  
[15]  
RP VARIANT CAIS MEM-866.  
RX MEDLINE=91186983; PubMed=2082179;  
RA Brown T.R., Lubahn D.B., Wilson E.M., French F.S., Migeon C.J.,  
RA Corfen J.L.;  
RT "Functional characterization of naturally occurring mutant androgen  
RT receptors from subjects with complete androgen insensitivity.";  
RL Mol. Endocrinol. 4:1759-1772(1990).  
[16]  
RP VARIANT CYS-774.  
RX MEDLINE=91310758; PubMed=1856263;  
RA Marceli M., Tilley W.D., Zoppi S., Griffin J.E., Wilson J.D.,  
RA McPhaul M.J.;  
RT "Androgen resistance associated with a mutation of the androgen  
RT receptor at amino acid 772 (Arg-->Cys) results from a combination of  
RT decreased messenger ribonucleic acid levels and impairment of  
RT receptor function.";  
RL J. Clin. Endocrinol. Metab. 73:318-325(1991).

[17] VARIANT CAIS ASN-695 AND HIS-695, AND SEQUENCE OF 629-723 FROM N.A.  
 MEDLINE=92131007; PubMed=1775137;  
 RA Ris-Stalpers C., Trilliro M.A., Kuiper G.G., Jenster G., Romalo G.,  
 Sai T., van Rooij H.C., Kaufman M., Rosenfield R.L., Jiao S.;  
 RT "Substitution of aspartic acid-686 by histidine or asparagine in the  
 human androgen receptor leads to a functionally inactive protein with  
 altered hormone-binding characteristics."; Mol. Endocrinol. 5:1562-1569(1991).  
 [18] VARIANTS CAIS AND PAIS.  
 MEDLINE=93338440; PubMed=1307250;  
 RA Batch J.A., Williams D.M., Davies H.R., Brown B.D., Evans B.A.J.,  
 Hughes I.A., Patterson M.N.;  
 RT "Androgen receptor gene mutations identified by SSCP in fourteen  
 subjects with androgen insensitivity syndrome."; Hum. Mol. Genet. 1:497-503(1992).  
 [19] VARIANT CAIS VAL-787.  
 MEDLINE=92235226; PubMed=1569153;  
 RA Nakao R., Hajj M., Yanase T., Ogo A., Takayanagi R., Katsube T.,  
 Fukumaki Y., Nawata H.;  
 RT "A single amino acid substitution (Met-786-->Val) in the steroid-  
 binding domain of human androgen receptor leads to complete androgen  
 insensitivity syndrome."; J. Clin. Endocrinol. Metab. 74:1152-1157(1992).  
 [20] VARIANT LNCAP ALA-877.  
 MEDLINE=92222955; PubMed=1562599;  
 RA Veldscholte J., Berrevoets C.A., Ris-Stalpers C., Kuiper G.G.,  
 Jenster G., Trapman J., Brinkman A.O., Mulder E.;  
 RT "The androgen receptor in LNCaP cells contains a mutation in the  
 ligand binding domain which affects steroid binding characteristics  
 and response to antiandrogens."; J. Steroid Biochem. Mol. Biol. 41:665-669(1992).  
 [21] VARIANT MET-730.  
 MEDLINE=92335289; PubMed=1631125;  
 RA Newmark J.R., Hardy D.O., Tonb D.C., Carter B.S., Epstein J.L.,  
 Isaacs W.B., Brown T.R., Barrack E.R.;  
 RT "Androgen receptor gene mutations in human prostate cancer."; Proc. Natl. Acad. Sci. U.S.A. 89:6319-6323(1992).  
 [22] VARIANT CAIS VAL-754.  
 MEDLINE=93372806; PubMed=8103398;  
 RA Lobbaccaro J.-M., Lombroso S., Ktari R., Dumas R., Sultan C.;  
 RT "An exonic point mutation creates a Maellit site in the androgen  
 receptor gene of a family with complete androgen Insensitivity  
 syndrome."; Hum. Mol. Genet. 2:1041-1043(1993).  
 [23] VARIANT CAIS ARG-807.  
 MEDLINE=94108430; PubMed=8281140;  
 RA Adegoke O., Kallio P.J., Palvimäki J.J., Kontula K., Jaenue O.A.;  
 RT "A single-base substitution in exon 6 of the androgen receptor gene  
 causing complete androgen insensitivity: the mutated receptor fails  
 to transactivate but binds to DNA in vitro."; Hum. Mol. Genet. 2:1809-1812(1993).  
 [24] VARIANT PAIS VAL-743.  
 MEDLINE=93315568; PubMed=8325932;  
 RA Nakao R., Yanase T., Sakai Y., Hajj M., Nawata H.;  
 RT "A single amino acid substitution (Gly743 --> Val) in the steroid-  
 binding domain of the human androgen receptor leads to Reflenstein

Query Match	22.3%	Score 79:	DB 1:	Length 919:
Best Local Similarity	32.8%	Pred No. 0.51:		
Matches 19, Conservative 15:		Mismatches 22:	Indels 2:	Gaps 1

QY 5 QQRREYEDCRRRCEDQEPROQHCCYLRCREQDRHGRCGDMNPRQGGSGRYEEGEEEO 62  
 ||::: :::: ||: | : ||:: | : ||: |  
 Db 58 QQQQQQ--QQQQQQQQQQQQQQQETSPRQQQQQQQGEDSGQAHRRCPRTGYLVLDEEQ 113

ID	NT4_NEUCR	STANDARD	PRT	1090 AA
AC	P28349.			
DT	01-DEC-1992 (Rel. 24, Created)			
DT	01-DEC-1992 (Rel. 24, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	NITROGEN ASSIMILATION TRANSCRIPTION FACTOR NIT-4.			
GN	NIT-4.			
OS	Neurospora crassa.			
OC	Eukaryota; Fungi; Ascomycota; Sordariales; Sordariaceae; Neurospora.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=92017855; PubMed=1840634;			
RA	Yuan G.-F., Fu Y.-H., Marzluf G.A.;			
RT	"nit-4, a pathway-specific regulatory gene of Neurospora crassa,			
RT	encodes a protein with a putative binuclear zinc DNA-binding			
RL	domain.";			
RL	Mol. Cell. Biol. 11:5735-5745(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=92149315; PubMed=1531376;			
RA	Yuan G.-F., Marzluf G.A.;			
RT	"Molecular characterization of mutations of nit-4, the			
RT	pathway-specific regulatory gene which controls nitrate assimilation			
RT	in Neurospora crassa.";			
RL	Mol. Microbiol. 6:67-73(1992).			
CC	-1 FUNCTION: PATHWAY-SPECIFIC REGULATORY GENE OF NITRATE			
CC	ASSIMILATION; IT ACTIVATES THE TRANSCRIPTION OF THE GENES FOR			
CC	NITRATE AND NITRITE REDUCTASES.			
CC	-1 SUBCELLULAR LOCATION: NUCLEAR.			
CC	-1 DOMAIN: THE GLUTAMINE-RICH DOMAIN MIGHT FUNCTION IN ACTIVATING			
CC	GENE EXPRESSION.			
CC	-1 SIMILARITY: CONTAINS A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR			
CC	CLUSTER DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; M80368; AAA33602.1; -			
DR	PIR; A41696; A41696.			
DR	PIR; S20033; S20033.			
DR	HSSP; P07272; 1PY1.			
DR	INTERPRO; IPR001138; -			
DR	PFAM; PF001172; zn-clus; 1.			
DR	PROSITE; PS00463; ZN2_CY6_FUNGAL_1; 1.			
DR	PROSITE; PS50048; ZN2_CY6_FUNGAL_2; 1.			
KW	Transcription regulation; Activator; DNA-binding; Nuclear protein;			
KW	Zinc; Metal-binding; Nitrate assimilation.			
FT	DNA_BIND	53	81	
FT	DOMAIN	121	139	
FT	DOMAIN	213	229	
FT	DOMAIN	429	450	
FT	DOMAIN	672	754	
FT	DOMAIN	755	859	
FT	DOMAIN	992	1024	
FT	CONFLICT	98	98	
FT	CONFLICT	467	467	
FT	SEQUENCE	1090 AA; 120244 MW; 881D89172EDD6114 CRC64;		

Query Match	21.1%	Score 75	DB 1	Length 1090
Best Local Similarity	32.7%	Pred. No. 1.3		
Matches 17	Conservative 14	Mismatches 17	Indels 4	Gaps 1
6 GREEDCRRRCEDQEPROHQHCOLRCREDDQOHGRGDDMM-----NPRGSSG 53				
11:	:	11:	11:	11:

6 QREYEDCRRCEQQEP RQHQ CQLRCREQQORQHGRGDM---NPQRGSG 53

```

Db          990   QKQDQQDQDQDQDQDQDQDQDQDQDQDQDEANMFATVHHAERGGG 1041

RESULT#    6
TRHY_HUMAN TRHY_HUMAN      STANDARD;             PRT;       1898 AA.
AC         007283:
DT         01-OCT-1994 (Rel. 30, Created)
DT         01-OCT-1994 (Rel. 30, Last sequence update)
DT         01-OCT-2000 (Rel. 40, Last annotation update)
DE         TRICHOXYALIN.
GN         TTH OR TRHY OR THL.
OS         Homo sapiens (Human).
OC         Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;
OC         Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
RN         [1]
RP         SEQUENCE FROM N.A.
RX         MEDLINE=933280194.; PubMed=7685034;
RA         Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,
RA         Steinert P.M.:
RT         "The structure of human trichohyalin, potential multiple roles as a
RT         functional EF-hand-like calcium-binding protein, a cornified cell
RT         envelope precursor, and an intermediate filament-associated (cross-
RT         linking) protein."
RL         J. Biol. Chem. 268:12164-12176(1993).
RN         (2)
RP         SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
RX         MEDLINE=93315897.; PubMed=7686553;
RA         O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinert P.M.:
RT         "Trichohyalin: a structural protein of hair, tongue, nail, and
RT         epidermis."
RL         J. Invest. Dermatol. 101:655-715(1993).
CC         -I- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC         IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC         INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC         LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC         ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC         WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC         ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC         ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC         DIFFERENTIATION.
CC         CC
CC         -I- SUBUNIT: MONOMER (PROBABLE).
CC         CC
CC         -I- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC         THE INNER ROOT SHEAT (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC         THE FLITIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
CC         CC
CC         -I- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
CC         THE EPIDERMIS.
CC         CC
CC         -I- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC         CALCULON-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC         ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC         OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STANDARD
CC         ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC         THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC         DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC         THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC         DIFFERENT SPECIES.
CC         CC
CC         -I- PM: KNOWN SUBSTRATE OF TRANSGUTAMINASE. SOME 200 ARGININES ARE
CC         PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.
CC         CC
CC         -I- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE S-100
CC         FAMILY.
CC         CC
CC         -I- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC         CC
CC         THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC         between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC         the European Bioinformatics Institute. There are no restrictions on its
CC         use by non-profit institutions as long as its content is in no way
CC         modified and this statement is not removed. Usage by and for commercial
CC         entities requires a license agreement (see http://www.isb-sib.ch/annouce/
CC         or send an email to license@isb-sib.ch).
CC         CC
RR         EMBL, L09190; AAA6582.1; -.
RR         PIR: A45973; A45973.
RR         HSSP: P02633.IBOC.
```

DR	MIM:	190370;	-
DR	INTERPRO:	IPR001751;	-
DR	INTERPRO:	IPR002048;	-
DR	Pfam:	PF01023;	S_100; 1.
DR	Pfam:	PF00036;	eHand; 1.
DR	PROSITE:	PS00018;	EF_HAND; 1.
DR	PROSITE:	PS00303;	S100_CABP; 1.
KW	Repeat:	Calcium-binding.	
FT	DOMAIN	1	91
FT	CA_BIND	22	33
FT	CA_BIND	62	73
FT	DOMAIN	314	390
FT	REPEAT	314	326
FT	REPEAT	327	339
FT	REPEAT	340	351
FT	REPEAT	352	364
FT	REPEAT	365	377
FT	REPEAT	378	390
FT	DOMAIN	391	444
FT	REPEAT	391	396
FT	REPEAT	397	402
FT	REPEAT	403	408
FT	REPEAT	409	414
FT	REPEAT	415	420
FT	REPEAT	421	426
FT	REPEAT	427	432
FT	REPEAT	433	438
FT	REPEAT	439	444
FT	DOMAIN	444	702
FT	DOMAIN	923	1162
FT	REPEAT	923	952
FT	REPEAT	953	982
FT	REPEAT	983	1012
FT	REPEAT	1013	1042
FT	REPEAT	1043	1072
FT	REPEAT	1073	1102
FT	REPEAT	1103	1132
FT	REPEAT	1133	1162
FT	DOMAIN	1250	1849
FT	CONFLICT	1752	1752
FT	CONFLICT	1794	1801
FT	CONFLICT	1857	1857
FT	CONFLICT	1880	1880
SEQ	SEQUENCE	1898 AA;	247219 MW; A74B5947FB62E31D CRC64;

Query Match 20.8%, Score 74; DB 1; Length 1898;  
 Best Local Similarity 36.4%; Pred. No. 3.1;  
 Matches 24; Conservative 10; Mismatches 28; Indels 4; Gaps 2;

```

OY      1 KRDPQQRVEEDCRRCCEGOQEPK-QQHOCQLKRREQQRHGKRGDMNPNQRGSGRYEGE 59
          :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db       1723 ERDKRFREEEOQLRGREEOQLRSQSDDKRFREEEOQLREREEQOLRPQQRDGKYWEEOQ 1782
          :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY      60 ---EEQ 62
          |||
Db       1783 LQLEEQ 1788
    
```

RESULT 7  
 CBP\_HUMAN STANDARD: PRT: 2442 AA.  
 AC Q92793; O15376; O00147;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE CREB-BINDING PROTEIN.  
 GN CREBBP OR CBP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 RN [1]





OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LYMPHOMA;  
 RA Fisher R., Fillmore H., Reynolds A.B.;  
 RT "Molecular cloning and characterization of the 162 kDa component  
 of a multi-protein complex phosphorylated by Src."  
 RT Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE OF 514-961 FROM N.A. (CENTROSOMIN B).  
 RA MEDLINE-98039715; PubMed-9372446;  
 RA Petzelt C., Joswig G., Mucheva A., Licher P., Stammer H., Werner D.;  
 RT "The centrosomal protein centrosomin A and the nuclear protein  
 centrosomin B derive from one gene by post-transcriptional processes  
 involving RNA editing."  
 RT J. Cell Sci. 110:2573-2578(1997).  
 RN (3)  
 RP SEQUENCE OF 514-790 FROM N.A. (CENTROSOMIN A).  
 RA MEDLINE-91277032; PubMed-1829085;  
 RA Joswig G., Petzelt C., Werner D.;  
 RT "Murine CDNA's coding for the centrosomal antigen centrosomin A."  
 RT J. Cell Sci. 98:37-43(1991).  
 RN (4)  
 RP REVISIONS.  
 RA Joswig G., Petzelt C., Werner D.;  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF  
 METHIONYL-TRNAI AND MRNA.  
 CC -1- SUBUNIT: EFF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
 CC -1- PTM: PHOSPHORYLATED.  
 CC -1- SIMILARITY: BELONGS TO THE EIF3S10 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: U14172; AAA09010.1; -  
 DR EMBL: X84651; CAA59144.1; -  
 DR EMBL: X17373; CAA55246.1; -  
 DR PIR: S13800; S13800.  
 DR MGD: MGI:95301; EIF3.  
 KW Initiation factor; Protein biosynthesis; Repeat; Phosphorylation.  
 FT DOMAIN 924 1133  
 FT 21 X 10 AA TANDM REPEAT OF D-[DE]-D-R-  
 FT CONFLICT 613 647  
 FT [GP]-[PS]-[RW]-R-[GN]-[AM].  
 FT ROEAKEREKRILOEHOIKKTVRELRQIKKTE ->  
 FT PRGKAREGERTNSRTFANGECGAVRADQEDR (IN  
 FT REF. 2 AND 3).  
 FT CONFLICT 683 684  
 FT EL -> DY (IN REF. 2 AND 3).  
 FT CONFLICT 717 717  
 FT Q -> H (IN REF. 2 AND 3).  
 FT CONFLICT 766 766  
 FT A -> V (IN REF. 2 AND 3).  
 FT CONFLICT 787 790  
 FT RHR -> SIVA (IN REF. 3).  
 FT CONFLICT 793 793  
 FT E -> D (IN REF. 2).  
 SO SEQUENCE 1344 AA; 161949 MW; F4CAE2169F577712 CRC64;

Query Match 19.9%; Score 70.5; DB 1; Length 1344;  
 Best Local Similarity 26.4%; Pred. No. 5.2;  
 Matches 19; Conservative 20; Mismatches 24; Indels 9; Gaps 3;

OY 1 KRDPOREVEDCRRRCEQDF-PRQCHQCLRCRQQRQGR--GGMMNPQ-----RGG 51  
 DB 836 KKEEELREYQERVKKLEVERKKRRELEIERERRERREERLRGDDPLSRKDSRWGDRDS 895  
 OY 52 SGRVEGEEES 63  
 DB 896 ESTWRKGRPEADS 907

RESULT 12  
 WT1\_SMIMA STANDARD; PRT; 239 AA.  
 AC P49953;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE WT1MS' TUMOR PROTEIN (FRAGMENT).  
 GN WT1.  
 OS Sinthopsis macronura (Stripe-faced dunart).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Metatheria; Dasyuromorpha; Dasyuridae; Sinthopsis.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-TESTIS;  
 RA MEDLINE-96068905; PubMed-7478606;  
 RA Kent J., Coriat A.M., Sharpe P.T., Hastie N., van Heyningen V.;  
 RT "The evolution of WT1 sequence and expression pattern in the  
 vertebrates."  
 RT Oncogene 11:1781-1792(1995).  
 CC -1- FUNCTION: POTENTIAL ROLE IN TRANSCRIPTIONAL REGULATION. RECOGNIZES  
 AND BINDS TO THE DNA SEQUENCE 5'-CGCCCCCC-3' (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- ALTERNATIVE PRODUCTS: TWO ALTERNATIVE SPLICE SITES EXISTS (BY  
 SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER  
 PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X85732; CAA59737.1; -  
 DR HSSP: P08046; IAL1.  
 DR INTERPRO: IPR000822; -  
 DR PIRAM: PR00096; zf-C2H2; 4.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2; 4.  
 KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;  
 KW Transcription regulation; Alternative splicing; Anti-oncogene.  
 FT NON\_TER 1 1  
 FT DOMAIN 113 228  
 FT ZINC\_FINGERS.  
 FT ZN\_FING 113 137  
 FT C2H2-TYPE.  
 FT ZN\_FING 143 167  
 FT C2H2-TYPE.  
 FT ZN\_FING 173 195  
 FT C2H2-TYPE.  
 FT ZN\_FING 204 228  
 FT C2H2-TYPE.  
 FT VARSPLIC 198 200  
 FT MISSING (IN ISOFORM 2).  
 SO SEQUENCE 239 AA; 27793 MW; 6707678A7259A624 CRC64;

Query Match 19.4%; Score 69; DB 1; Length 239;  
 Best Local Similarity 25.0%; Pred. No. 1.5;  
 Matches 15; Conservative 14; Mismatches 25; Indels 6; Gaps 1;

OY 4 PQOREVEDCRRRCEQDFPRQCHQ-----LRCRQQRQHGGRGMMNPQRCGSGRYEE 57  
 DB 142 PYOCDEKDCRRRSRDLKRHQRHTTGAVKPPCKCKRFSKSHLKTHTHTGKTS 201  
 RESULT 13  
 WT1\_ALIMI STANDARD; PRT; 288 AA.  
 ID WT1\_ALIMI  
 AC P50902;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE WT1MS' TUMOR PROTEIN (FRAGMENT).  
 GN WT1.



```

OS Alligator mississippiensis (American alligator).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Atherosauria; Crocodylidae; Alligatorinae; Alligator.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96068905; Pubmed=7478606;
RA Kent J., Coriat A.M., Sharpe P.T., Hastie N., van Heyningen V.;
RT "The evolution of Wt1 sequence and expression pattern in the
vertebrates."
RL Oncogene 11:1781-1792(1995).
CC -1- FUNCTION: POTENTIAL ROLE IN TRANSCRIPTIONAL REGULATION. RECOGNIZES
CC AND BINDS TO THE DNA SEQUENCE 5'-CGCCCCCG-3'.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: TWO ALTERNATIVE SPLICED SITES EXISTS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X85730; CAA59735.1; -.
DR HSSP; P08046; 1A1L.
DR INTERPRO; IPR000822; -.
DR PRAM; PF00096; zf-C2H2; 4.
DR PROSITE; PS00028; ZINC_FINGER_C2H2; 4.
KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KW Transcription regulation; Alternative splicing.
FT NCN_TER 1
FT ZN_FING 162 186 C2H2-TYPE.
FT ZN_FING 192 216 C2H2-TYPE.
FT ZN_FING 222 244 C2H2-TYPE.
FT ZN_FING 253 277 C2H2-TYPE.
FT VARSPIC 245 247 MISSING (IN ISOFORM 2).
SO SEQUENCE 288 AA; 33111 MW; 33E267DBE7BFDE CRC64;

Query Match 19.4%; Score 69; DB 1; Length 288;
Best Local Similarity 25.0%; Pred. No. 1.8;
Matches 15; Conservative 14; Mismatches 25; Indels 6; Gaps 1;

OY 4 POOREYEDCRRCRCEQZPRROHQO-----LRCREOQRHGRCGDMNQRGSGGYEE 57
DB 191 PYOCDFDCRRRSRSDQLKRHRHTGVKRPCKTCQRKFSRSDHLKTRHTHTGKTS 250

RESULT 14
WT1_RAT
ID WT1_RAT STANDARD; PRT; 448 AA.
AC P49562;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE WILMS' TUMOR PROTEIN HOMOLOG.
GN WT1 OR WT-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=SPRAGUE-DAWLEY; TISSUE=KIDNEY;
RX MEDLINE=93046155; Pubmed=1330293;
RA Sharma P.M., Yang X., Bowman M., Roberts V., Sukumar S.;
RT "Molecular cloning of rat Wilms' tumor complementary DNA and a study
RT of messenger RNA expression in the urogenital system and the brain."
RL Cancer Res. 52:6407-6412(1992).
CC -1- FUNCTION: POTENTIAL ROLE IN TRANSCRIPTIONAL REGULATION. RECOGNIZES
CC AND BINDS TO THE DNA SEQUENCE 5'-CGCCCCCG-3'.

```

```

CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- ALTERNATIVE PRODUCTS: TWO ALTERNATIVE SPLICED SITES EXISTS.
CC -1- TISSUE SPECIFICITY: KIDNEY (BY SIMILARITY).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING KIDNEY DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X69716; CAA49373.1; -.
DR TRANSFAC; T02352; -.
DR INTERPRO; IPR000822; -.
DR PRAM; PF00096; zf-C2H2; 4.
DR PRINTS; PR00048; ZINC_FINGER.
DR PROSITE; PS00049; WILMSTUMOR.
KW Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
KW Transcription regulation; Alternative splicing; Anti-oncogene.
FT DOMAIN 27 82
FT ZINC_FINGERS.
FT ZN_FING 322 437
FT ZN_FING 352 346 C2H2-TYPE.
FT ZN_FING 372 376 C2H2-TYPE.
FT ZN_FING 382 404 C2H2-TYPE.
FT ZN_FING 413 437 C2H2-TYPE.
FT VARSPIC 249 265 MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT VARSPIC 407 409 MISSING (IN ISOFORM 2 AND ISOFORM 4).
SO SEQUENCE 448 AA; 49193 MW; 329AC9AC1FF73F76 CRC64;

Query Match 19.4%; Score 69; DB 1; Length 448;
Best Local Similarity 25.0%; Pred. No. 2.7;
Matches 15; Conservative 14; Mismatches 25; Indels 6; Gaps 1;

OY 4 POOREYEDCRRCRCEQZPRROHQO-----LRCREOQRHGRCGDMNQRGSGGYEE 57
DB 351 PYOCDFDCRRRSRSDQLKRHRHTGVKRPCKTCQRKFSRSDHLKTRHTHTGKTS 410

RESULT 15
WT1_HUMAN
ID WT1_HUMAN STANDARD; PRT; 449 AA.
AC P19544;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE WILMS' TUMOR PROTEIN (WT33).
GN WT1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=FETAL KIDNEY;
RX MEDLINE=90158822; Pubmed=2154702;
RA Gessler M., Poustka A., Cavenee W., Neve R.L., Orkin S.H.,
RA Bruns G.A.P.;
RT "Homologous deletion in Wilms tumours of a zinc-finger gene
RT identified by chromosome jumping."
RL Nature 343:774-778(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=PLACENTA;
RX MEDLINE=92052142; Pubmed=1658787;
RA Haber D.A., Sohn R.L., Buckler A.J., Pelletier J., Call K.M.,
RA Housman D.E.;
RT "Alternative splicing and genomic structure of the Wilms tumor gene

```

RT WTI.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:9618-9622(1991).  
 RN [3]  
 RX SEQUENCE OF 85-449 FROM N.A.  
 RP MEDLINE=90150277; PubMed=2154335;  
 RA Call K.M., Glaser T., Ito C.Y., Buckler A.J., Pelletier J.,  
 RA Haber D.A., Rose E.A., Kral A., Yeger H., Lewis W.H., Jones C.,  
 RA Housman D.E.;  
 RT "Isolation and characterization of a zinc finger polypeptide gene at  
 RT the human chromosome 11 Wilms' tumor locus.";  
 RL Cell 60:509-520(1990).  
 RN [4]  
 RP IDENTIFICATION OF START CODON AND ALTERNATIVE SPLICE SITES.  
 RX MEDLINE=91141522; PubMed=1671709;  
 RA Buckler A.J., Pelletier J., Haber D.A., Glaser T., Housman D.E.;  
 RT "Isolation, characterization, and expression of the murine Wt1"  
 RT tumor gene (WT1) during kidney development.";  
 RL Mol. Cell. Biol. 11:1707-1712(1991).  
 RN [5]  
 RP VARIANT WT CYS-366.  
 RX MEDLINE=92279213; PubMed=1317572;  
 RA Little M.H., Prosser J., Condie A., Smith P.J., van Heyningen V.,  
 RA Hastie N.D.;  
 RT "Zinc finger point mutations within the WT1 gene in Wilms tumor  
 RT patients.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:4791-4795(1992).  
 RN [6]  
 RP VARIANTS DDS.  
 RX MEDLINE=92005721; PubMed=1655284;  
 RA Pelletier J., Bruenling W., Kashan C.E., Mauer S.M., Manivel J.C.,  
 RA Striegel J.E., Houghton D.C., Junien C., Habib R., Fouser L.,  
 RA Fine R.N., Silverman B.L., Haber D.A., Housman D.;  
 RT "Germline mutations in the Wt1 tumor suppressor gene are  
 RT associated with abnormal urogenital development in Denys-Drash  
 RT syndrome.";  
 RL Cell 67:437-447(1991).  
 RN [7]  
 RP VARIANTS DDS.  
 RX MEDLINE=93265053; PubMed=1338906;  
 RA Baird P.N., Santos A., Groves N., Jadresic L., Cowell J.K.;  
 RT "Constitutional mutations in the WT1 gene in patients with  
 RT Denys-Drash syndrome.";  
 RL Hum. Mol. Genet. 1:301-305(1992).  
 RN [8]  
 RP VARIANTS DDS.  
 RX MEDLINE=93271983; PubMed=8388765;  
 RA Little M.H., Williamson K.A., Mannens M., Kelsey A., Gosden C.,  
 RA Hastie N., van Heyningen V.;  
 RT "Evidence that WT1 mutations in Denys-Drash syndrome patients may act  
 RT in a dominant-negative fashion.";  
 RL Hum. Mol. Genet. 2:259-264(1993).  
 RN [9]  
 RP VARIANT MESOTHELIOMA GLY-273.  
 RX MEDLINE=94004972; PubMed=8401592;  
 RA Park S., Schalling M., Bernard A., Maheswaran S., Shipley G.C.,  
 RA Roberts D., Fletcher J., Shipman R., Rheinwald J., Demetri G.,  
 RA Griffin J., Minden M., Housman D.E., Haber D.A.;  
 RT "The Wt1s tumor gene WT1 is expressed in murine mesoderm-derived  
 RT tissues and mutated in a human mesothelioma.";  
 RL Nat. Genet. 4:415-420(1993).  
 RN [10]  
 RP VARIANTS WT SER-181 AND ALA-253.  
 RX MEDLINE=97268681; PubMed=9108089;  
 RA Schumacher V., Schneider S., Figge A., Wildhardt G., Harms D.,  
 RA Schmidt D., Weirich A., Ludwig R., Royer-Pokora B.;  
 RT "Correlation of germ-line mutations and two-hit inactivation of the  
 RT WT1 gene with Wilms tumors of stromal-predominant histology.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:3972-3977(1997).  
 RN [11]  
 RP VARIANTS DMS TYR-377; LEU-383 AND ASN-396.  
 RX MEDLINE=98198341; PubMed=9529364;  
 RA Jeanpierre C., Denamur E., Henry I., Cabanis M.-O., Luce S.,  
 RA Cecille A., Elion J., Peuchmaur M., Lohrat C., Naudet P.,

RA Gubler M.-C., Junien C.;  
 RT "Identification of constitutional WT1 mutations, in patients with  
 RT isolated diffuse mesangial sclerosis, and analysis of  
 RT genotype/phenotype correlations by use of a computerized mutation  
 RT database.";  
 RL Am. J. Hum. Genet. 62:824-833(1998).  
 RN [12]  
 RP REVIEW.  
 RX MEDLINE=92207913; PubMed=1313285;  
 RA Haber D.A., Buckler A.J.;  
 RT "WT1: a novel tumor suppressor gene inactivated in Wilms' tumor.";  
 RL New Biol. 4:97-106(1992).  
 RN [13]  
 RP REVIEW.  
 RX MEDLINE=93345769; PubMed=8393820;  
 RA Rauscher F.J. III;  
 RT "The WT1 Wilms tumor gene product: a developmentally regulated  
 RT transcription factor in the kidney that functions as a tumor  
 RT suppressor.";  
 RL FASEB J. 7:896-903(1993).  
 CC -1- FUNCTION: POTENTIAL ROLE IN TRANSCRIPTIONAL REGULATION. RECOGNIZES  
 CC AND BINDS TO THE DNA SEQUENCE 5'-CGCCCCCC-3'.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- ALTERNATIVE PRODUCTS: TWO ALTERNATIVE SPLICE SITES EXISTS.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE KIDNEY AND A SUBSET OF  
 CC HEMATOPOIETIC CELLS.  
 CC -1- DISEASE: WILM'S TUMOR (WT) IS AN EMBRYONAL MALIGNANCY OF THE  
 CC KIDNEY THAT AFFECTS APPROXIMATELY 1 IN 10,000 INFANTS AND YOUNG  
 CC CHILDREN. IT OCCURS BOTH IN SPORADIC AND HEREDITARY FORMS.  
 CC -1- INACTIVATION OF WT1 IS ONE OF THE CAUSES OF WILM'S TUMOR.  
 CC -1- DISEASE: PATIENTS WITH DENYS-DRASH SYNDROME (DDS), WHICH IS  
 CC CHARACTERIZED BY A TYPICAL NEPHROPATHY AND GENITAL ABNORMALITIES,  
 CC HAVE DEFECTS IN THE WT1 GENE.  
 CC -1- DISEASE: DEFECTS IN WT1 ARE ALSO A CAUSE OF DIFFUSE MESANGIAL  
 CC SCLEROSIS (DMS), A FORM A DDS.  
 CC -1- SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER  
 CC PROTEINS.  
 CC -1- DATABASE: NAME=WT1; NOTE=WT1 mutation database;  
 CC WWW="http://www.umd.necker.fr:2003/".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X51630; CAA35956.1; ALT\_INIT.  
 DR EMBL: M80232; AAA61299.1; -  
 DR EMBL: M80217; AAA61299.1; JOINED.  
 DR EMBL: M80218; AAA61299.1; JOINED.  
 DR EMBL: M80219; AAA61299.1; JOINED.  
 DR EMBL: M80220; AAA61299.1; JOINED.  
 DR EMBL: M80221; AAA61299.1; JOINED.  
 DR EMBL: M80228; AAA61299.1; JOINED.  
 DR EMBL: M80229; AAA61299.1; JOINED.  
 DR EMBL: M80231; AAA61299.1; JOINED.  
 DR EMBL: M30393; AAA36810.1; -  
 DR PIR: A34673; A34673.  
 DR PIR: S08273; S08273.  
 DR TRNSFAC: T00899; -  
 DR MIM: 194070; -  
 DR MIM: 194080; -  
 DR MIM: 256370; -  
 DR INTERPRO: IPR000822; -  
 DR INTERPRO: IPR000976; -  
 DR PEFAM: PFO00096; zF-C2H2; 4.  
 DR PRINTS: PR00048; ZINCINGER.  
 DR PRINTS: PR00049; WILMSTUMOUR.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2; 4.  
 DR Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;  
 KW Transcription regulation; Alternative splicing; Anti-oncogene;

```
KW Disease mutation.
FT DOMAIN 27 83 PRO-RICH.
FT DOMAIN 323 438 ZINC FINGERS.
FT ZN_FING 323 347 C2H2-TYPE.
FT ZN_FING 353 377 C2H2-TYPE.
FT ZN_FING 383 405 C2H2-TYPE.
FT ZN_FING 414 438 C2H2-TYPE.
FT VARSPLIC 250 266 MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT VARSPLIC 408 410 MISSING (IN ISOFORM 2 AND ISOFORM 4).
FT VARIANT 181 181 P -> S (IN WT).
FT VARIANT 223 223 /FTId=VAR_007739.
FT VARIANT 223 223 S -> N (IN WT).
FT VARIANT 253 253 /FTId=VAR_007740.
FT VARIANT 253 253 G -> A (IN WT).
FT VARIANT 273 273 /FTId=VAR_007741.
FT VARIANT 273 273 S -> G (IN MESOTHELIOMA).
FT VARIANT 330 330 /FTId=VAR_007742.
FT VARIANT 330 330 C -> Y (IN DDS).
FT VARIANT 360 360 /FTId=VAR_007743.
FT VARIANT 360 360 C -> G (IN DDS).
FT VARIANT 366 366 /FTId=VAR_007744.
FT VARIANT 366 366 R -> C (IN WT).
FT VARIANT 366 366 /FTId=VAR_007745.
FT VARIANT 366 366 R -> H (IN DDS).
FT VARIANT 373 373 /FTId=VAR_007746.
FT VARIANT 373 373 H -> Q (IN DDS).
FT VARIANT 377 377 /FTId=VAR_007747.
FT VARIANT 377 377 H -> Y (IN DDS/DMS).

Query Match 19.48; Score 69; DB 1; Length 449;
Best Local Similarity 25.08; Pred. No. 2.7;
Matches 15; Conservative 14; Mismatches 25; Indels 6; Gaps 1; .

QY 4 POREVEDCRRCRCEQREPROQHOC-----LRCREOQRHGRGDMNPQRGGSGRYEE 57
Db 352 PYQDFRDCERRSRSDQLKRHRHTGVKPFQCKTCQRRKFSRSDHLKTYHTRHTGKTSE 411
```

Search completed: March 1, 2001, 16:16:33  
Job time: 413 sec

